Learning Spatially Localized, Parts-Based Representation

Stan Z. Li[1], XinWen Hou[2]*, HongJiang Zhang[1], QianSheng Cheng[2].
[1] Microsoft Research China, Beijing Sigma Center, Beijing 100080, China
[2] Institute of Mathematical Sciences, Peking University, Beijing, China
Contact: szli@microsoft.com, http://research.microsoft.com/∼szli

Abstract

In this paper, we propose a novel method, called local non-negative matrix factorization (LNMF), for learning spatially localized, parts-based subspace representation of visual patterns. An objective function is defined to impose localization constraint, in addition to the non-negativity constraint in the standard NMF [1]. This gives a set of bases which not only allows a non-subtractive (part-based) representation of images but also manifests localized features. An algorithm is presented for the learning of such basis components. Experimental results are presented to compare LNMF with the NMF and PCA methods for face representation and recognition, which demonstrates advantages of LNMF.

1 Introduction

Subspace analysis helps to reveal low dimensional structures of patterns observed in high dimensional spaces. A specific pattern of interest can reside in a low dimensional sub-manifold in the original input data space of an unnecessarily high dimensionality. Consider the case of $N \times M$ image pixels, each taking a value in $\{0, 1, \ldots, 255\}$; there is a huge number of possible configurations: $256^{N \times M}$. This space is capable of describing a wide variety of patterns or visual object classes. However, for a specific pattern, such as the human face, the number of admissible configurations is a only tiny fraction of that. In other words, the intrinsic dimension is much lower than $N \times M$.

An observation can be considered as a consequence of linear or nonlinear fusion of a small number of hidden or latent variables. Subspace analysis is aimed to derive a representation that results in such a fusion. In fact, the essence of feature extraction in pattern analysis can be considered as discovering and computing intrinsic low dimension of the pattern from the observation.

For these reasons, subspace analysis has been a major research issue in appearance based imaging and vision, such as object detection and recognition [2, 3, 4, 5, 6, 7]. The significance of is twofold: effective characterization of the pattern and dimension reduction.

One approach for learning a subspace representation for a class of image patterns involves deriving a set of basis components for construction of the subspace. The eigeniamge method [2, 3, 4] uses principal component analysis (PCA) [8] performed on a set of representative training data to decorrelate second order moments corresponding to low frequency property. Any image can be represented as a linear combination of these bases. Dimension reduction is achieved by discarding least significant components. Due to the holistic nature of the method, the resulting components are global interpretations, and thus PCA is unable to extract basis components manifesting localized features.

However, in many applications, localized features offer advantages in object recognition, including stability to local deformations, lighting variations, and partial occlusion. Several methods have been proposed recently for localized (spatially), parts-based (non-subtractive) feature extraction.

Local feature analysis (LFA) [9], also based on second order statistics, is a method for extracting, from the holistic (global) PCA basis, local topographic representation in terms of local features. Independent component analysis [10, 11] is a linear non-orthogonal transform. It yields a representation in which unknown linear mixtures of multi-dimensional random variables are made as statistically independent as possible. It not only decorrelates the second order statistics but also reduces higher-order statistical dependencies. It is found that independent component of natural scenes are localized edge-like filters [12].

The projection coefficients for the linear combinations in the above methods can be either positive or negative, and such linear combinations generally involve complex cancellations between positive and negative numbers. Therefore, these representations lack the intuitive meaning of adding parts to form a whole.

Non-negative matrix factorization (NMF) [1] imposes the non-negativity constraints in learning basis images. The pixel values of resulting basis images, as well as coefficients for reconstruction, are all non-negative. This way,

---

*The work presented in the paper was carried out at Microsoft Research, China.

only non-subtractive combinations are allowed. This ensures that the components are combined to form a whole in the non-subtractive way. For this reason, NMF is considered as a procedure for learning a parts-based representation [1]. However, the additive parts learned by NMF are not necessarily localized, and moreover, we found that the original NMF representation yields low recognition accuracy, as will be shown.

In this paper, we propose a novel subspace method, called local non-negative matrix factorization (LNMF), for learning spatially localized, parts-based representation of visual patterns. Inspired by the original NMF [1], the aim of this work to impose the locality of features in basis components and to make the representation suitable for tasks where feature localization is important. An objective function is defined to impose the localization constraint, in addition to the non-negativity constraint of [1]. A procedure is presented to optimize the objective to learn truly localized, parts-based components. A proof of the convergence of the algorithm is provided.

The rest of the paper is organized as follows: Section 2 introduces NMF in contrast to PCA. This is followed by the formulation of LNMF. A LNMF learning procedure is presented and its convergence proved. Section 3 presents experimental results illustrating properties of LNMF and its performance in face recognition as compared to PCA and NMF.

2 Constrained Non-Negative Matrix Factorization

Let a set of N_T training images be given as an $n \times N_T$ matrix $\mathbf{X} = [x_{ij}]$, with each column consisting of the n non-negative pixel values of an image. Denote a set of $m \leq n$ basis images by an $n \times m$ matrix $\mathbf{B}$. Each image can be represented as a linear combination of the basis images using the approximate factorization

$$\mathbf{X} \approx \mathbf{BH} \qquad (1)$$

where $\mathbf{H}$ is the matrix of $m \times N_T$ coefficients or weights. Dimension reduction is achieved when $m < n$.

The PCA factorization requires that the basis images (columns of $\mathbf{B}$ be orthonormal and the rows of $\mathbf{H}$ be mutually orthogonal. It imposes no other constraints than the orthogonality, and hence allows the entries of $\mathbf{B}$ and $\mathbf{H}$ to be of arbitrary sign. Many basis images, or eigenfaces in the case of face recognition, lack intuitive meaning, and a linear combination of the bases generally involves complex cancellations between positive and negative numbers. The NMF and LNMF representations allow only positive coefficients and thus non-subtractive combinations.

2.1 NMF

NMF imposes the non-negativity constraints instead of the orthogonality. As the consequence, the entries of $\mathbf{b}$ and $\mathbf{h}$ are all non-negative, and hence only non-subtractive combinations are allowed. This is believed to be compatible to the intuitive notion of combining parts to form a whole, and is how NMF learns a parts-based representation [1]. It is also consistent with the physiological fact that the firing rate are non-negative.

NMF uses the divergence of $\mathbf{X}$ from $\mathbf{Y}$, defined as

$$D(\mathbf{X}||\mathbf{Y}) = \sum_{i,j} \left(x_{ij} \log \frac{x_{ij}}{y_{ij}} - x_{ij} + y_{ij} \right) \qquad (2)$$

as the measure of cost for factorizing $\mathbf{X}$ into $\mathbf{BH} \stackrel{\triangle}{=} \mathbf{Y} = [y_{ij}]$. An NMF factorization is defined as

$$\min_{\mathbf{B},\mathbf{H}} \quad D(\mathbf{X}||\mathbf{BH}) \qquad (3)$$
$$\text{s.t} \quad \mathbf{B}, \mathbf{H} \geq \mathbf{0}, \ \sum_i b_{ij} = 1 \ \forall j$$

where $\mathbf{B}, \mathbf{H} \geq \mathbf{0}$ means that all entries of $\mathbf{B}$ and $\mathbf{H}$ are non-negative. $D(\mathbf{X}||\mathbf{Y})$ reduces to Kullback-Leibler divergence when $\sum_{i,j} x_{ij} = \sum_{i,j} y_{ij} = 1$. The above optimization can be done by using multiplicative update rules [13], for which a matlab program is available at `http://journalclub.mit.edu` under the "Computational Neuroscience" discussion category.

2.2 LNMF

The NMF model defined by the constrained minimization of (2) does not impose any constraints on the spatial locality and therefore minimizing the objective function can hardly yield a factorization which reveals local features in the data $\mathbf{X}$. Letting $\mathbf{U} = [u_{ij}] = \mathbf{B}^T\mathbf{B}$, $\mathbf{V} = [v_{ij}] = \mathbf{HH}^T$, both being $m \times m$, LNMF is aimed at learning local features by imposing the following three additional constraints on the NMF basis:

1. A basis component should not be further decomposed into more components, so as to minimize the number of basis components required to represent $\mathbf{X}$. Let $\mathbf{b}_j = [b_{ij}]_{i=1}^n$ be a basis vector. Given the existing constraints $\sum_i b_{ij} = 1$ for all i, we wish that $\sum_i b_{ij}^2$ should be as small as possible so that $\mathbf{b}_i$ contains as many non-zero elements as possible. This can be imposed by $\sum_i u_{ii} = \min$.

2. Different bases should be as orthogonal as possible, so as to minimize redundancy between different bases. This can be imposed by $\sum_{i \neq j} u_{ij} = \min$.

2

3. Only components giving most important information should be retained. Given that every image in $\mathbf{X}$ has been normalized into a certain range such as $[0, \ldots, 255]$, the total "activity" on each retained component, defined as the total squared projection coefficients summed over all training images, should be maximized. This is imposed by $\sum_i v_{ii} = \max$.

The incorporated of the above constraints leads the following constrained divergence as the objective function for LNMF:

$$D(\mathbf{X}||\mathbf{BH}) = \tag{4}$$
$$\sum_{i,j} \left(x_{ij} \log \frac{x_{ij}}{y_{ij}} - x_{ij} + y_{ij} \right) +$$
$$\alpha \sum_{i,j} u_{ij} - \beta \sum_i v_{ii}$$

where $\alpha, \beta > 0$ are some constants. A LNMF factorization is defined as a solution to the constrained minimization of (5).

A local solution to the above constrained minimization can be found by using the following three step update rules:

$$h_{kl} = \sqrt{h_{kl} \sum_i x_{il} \frac{b_{ik}}{\sum_k b_{ik} h_{kl}}} \tag{5}$$

$$b_{kl} = \frac{b_{kl} \sum_j x_{kj} \frac{h_{lj}}{\sum_k b_{kl} h_{lj}}}{\sum_j h_{lj}} \tag{6}$$

$$b_{kl} = \frac{b_{kl}}{\sum_k b_{kl}} \tag{7}$$

2.3 Convergence Proof

The learning algorithm (5)-(7) alternates between updating $\mathbf{H}$ and updating $\mathbf{B}$, which is derived based on a technique in which an objective function $L(\mathbf{Z})$ is minimized by using an auxiliary function. $G(\mathbf{H}, \mathbf{Z}')$ is said to be an auxiliary function for $L(\mathbf{Z})$ if $G(\mathbf{Z}, \mathbf{Z}') \geq L(\mathbf{Z})$ and $G(\mathbf{Z}, \mathbf{Z}) = L(\mathbf{Z})$ are satisfied. If G is an auxiliary function, then $L(\mathbf{Z})$ is non-increasing when $\mathbf{Z}$ is updated using $\mathbf{Z}^{(t+1)} = \arg\min_{\mathbf{Z}} G(\mathbf{Z}, \mathbf{Z}^{(t)})$ [14]. This is because $L(\mathbf{Z}^{(t+1)}) \leq G(\mathbf{Z}^{(t+1)}, \mathbf{Z}^{(t)}) \leq G(\mathbf{Z}^{(t)}, \mathbf{Z}^{(t)}) = L(\mathbf{Z}^{(t)})$.

Updating $\mathbf{H}$: $\mathbf{H}$ is updated by minimizing $L(\mathbf{H}) = \overline{D(\mathbf{X}||\mathbf{BH})}$ with $\mathbf{B}$ fixed. An auxiliary function is constructed for $L(\mathbf{H})$ as

$$G(\mathbf{H}, \mathbf{H}') = \tag{8}$$
$$\sum_{i,j} x_{ij} \log x_{ij} -$$
$$\sum_{i,j,k} x_{ij} \frac{b_{ik} h'_{kj}}{\sum_k b_{ik} h'_{kj}} \left(\log(b_{ik} h_{kj}) - \log \frac{b_{ik} h'_{kj}}{\sum_k b_{ik} h'_{kj}} \right) +$$

$$\sum_{i,j} y_{ij} - \sum_{i,j} x_{ij} + \alpha \sum_{i,j} u_{ij} - \beta \sum_{i,j} v_{ij}$$

It is easy to verify $G(\mathbf{H}, \mathbf{H}) = L(\mathbf{H})$. The following proves $G(\mathbf{H}, \mathbf{H}') \geq L(\mathbf{H})$. Because $\log\left(\sum_k b_{ik} h_{kj}\right)$ is a convex function, the following holds for all i, j and $\sum_k \mu_{ijk} = 1$:

$$-\log\left(\sum_k b_{ik} h_{kj}\right) \leq -\sum_k \mu_{ijk} \log \frac{b_{ik} h_{kj}}{\gamma \mu_{ijk}} \tag{9}$$

Let

$$\mu_{ijk} = \frac{b_{ik} h'_{kj}}{\sum_k b_{ik} h'_{kj}} \tag{10}$$

Then

$$-\log\left(\sum_k b_{ik} h_{kj}\right) \leq \tag{11}$$
$$-\sum_k \frac{b_{ik} h'_{kj}}{\sum_k b_{ik} h'_{kj}} \left(\log b_{ik} h_{kj} - \log \frac{b_{ik} h'_{kj}}{\sum_k b_{ik} h'_{kj}} \right)$$

which is $G(\mathbf{H}, \mathbf{H}') \geq L(\mathbf{H})$.

To minimize $L(\mathbf{H})$ w.r.t. $\mathbf{H}$, we can update $\mathbf{H}$ using

$$\mathbf{H}^{(t+1)} = \arg\min_{\mathbf{H}} G(\mathbf{H}, \mathbf{H}^{(t)}) \tag{12}$$

Such an $\mathbf{H}$ can be found by letting $\frac{\partial G(\mathbf{H}, \mathbf{H}')}{\partial h_{kl}} = 0$ for all kl. Because

$$\frac{\partial G(\mathbf{H}, \mathbf{H}')}{\partial h_{kl}} = -\sum_i x_{il} \frac{b_{ik} h'_{kl}}{\sum_k b_{ik} h'_{kl}} \frac{1}{h_{kl}} + \sum_i b_{ik} - 2\beta h_{kl} \tag{13}$$

we find

$$h_{kl} = \frac{1 - \sqrt{1 - 8\beta \sum_i x_{il} \frac{b_{ik} h'_{kl}}{\sum_k b_{ik} h'_{kl}}}}{4\beta} \tag{14}$$

There exists γ such that

$$h_{kl} \approx \gamma \sqrt{\sum_i x_{il} \frac{b_{ik} h'_{kl}}{\sum_k b_{ik} h'_{kl}}} \triangleq \gamma \tilde{h}_{kl} \tag{15}$$

where

$$\tilde{h}_{kl} = \sqrt{h'_{kl} \sum_i x_{il} \frac{b_{ik}}{\sum_k b_{ik} h'_{kl}}} \tag{16}$$

and $\gamma = \gamma(\beta, w)$ is a function of β and $w = \sum_i x_{il} \frac{b_{ik} h'_{kl}}{\sum_k b_{ik} h'_{kl}}$.

The result we want to derive from the LNMF learning is the basis $\mathbf{B}$, and $\mathbf{H}$ itself is not so important. Because b_{kl} will be normalized by Eq.7, the normalized b_{kl} is regardless of the γ value as long as $\gamma > 0$. Therefore, we simply replace Eq.(15) by (5).

3

Updating B: **B** is updated by minimizing $L(\mathbf{B}) = D(\mathbf{X}\|\mathbf{BH})$ with **H** fixed. The auxiliary function for $L(\mathbf{B})$ is

$$G(\mathbf{B}, \mathbf{B}') = \qquad\qquad (17)$$
$$\sum_{i,j} x_{ij} \log x_{ij} -$$
$$\sum_{i,j,k} x_{ij} \frac{b'_{ik} h_{kj}}{\sum_k b'_{ik} h_{kj}} \left(\log(b_{ik} h_{kj}) - \log \frac{b'_{ik} h_{kj}}{\sum_k b'_{ik} h_{kj}} \right) +$$
$$\sum_{i,j} y_{ij} - \sum_{i,j} x_{ij} + \alpha \sum_{i,j} u_{ij} - \beta \sum_{i,j} v_{ij}$$

We can prove $G(\mathbf{B}, \mathbf{B}) = L(\mathbf{B})$ and $G(\mathbf{B}, \mathbf{B}') \geq L(\mathbf{B})$ likewise. By letting $\frac{\partial G(\mathbf{B}, \mathbf{B}')}{\partial h_{kl}} = 0$, we find

$$b_{kl} = \frac{b'_{kl} \sum_j x_{kj} \frac{h_{lj}}{\sum_l b'_{kl} h_{lj}}}{\sum_j h_{lj} + 2\alpha \sum_j b_{kj}} = \frac{b'_{kl} \sum_j x_{kj} \frac{h_{lj}}{\sum_l b'_{kl} h_{lj}}}{\gamma \left(\sum_j \tilde{h}_{lj} + 2\frac{\alpha}{\gamma} \sum_j b_{kj} \right)} \quad (18)$$

Because $b_{ij} \in [0,1]$ and **B** is an approximately orthogonal basis, and $x_{ij} \in [0,255]$, there must be $h_{ij} \geq x_{ij} \in [0,255]$. Therefore, we can always set the ratio $\frac{\alpha}{\gamma}$ to be not too large (*e.g.* $\frac{\alpha}{\gamma} \approx 1$) so that $\sum_j \tilde{h}_{lj} + 2\frac{\alpha}{\gamma} \sum_j b_{kj}$ and thus $\tilde{h}_{lj} \gg 2\frac{\alpha}{\gamma} \sum_j b_{kj} \approx \tilde{h}_{lj}$. Therefore, we have Eq.(6).

From the above analysis, we conclude that the three step update rules (5)-(7) results in a sequence of non-increasing values of $D(\mathbf{X}|\mathbf{BH})$, and hence converges to a local minimum of it.

2.4 Face Recognition in Subspace

Face recognition in the PCA, NMF or LNMF linear subspace is performed as follows:

1. Feature extraction. Let $\bar{\mathbf{x}}$ be the mean of training images. Each training face image $\mathbf{x}_i$ is projected into the linear space as a feature vector $\mathbf{h}_i = \mathbf{B}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})$ which is then used as a prototype feature point. A query face image $\mathbf{q}$ to be classified is represented by its projection into the space as $\mathbf{h}_q = \mathbf{B}^{-1}(\mathbf{q} - \bar{\mathbf{x}})$.

2. Nearest neighbor classification. The Euclidean distance between the query and each prototype, $d(\mathbf{h}_q, \mathbf{h}_i)$, is calculated. The query is classified to the class to which the closest prototype belongs.

3 Experiments

3.1 Data Preparation

The Cambridge ORL face database is used for deriving PCA, NMF and LNMF bases. There are 400 images

(112×92) of 40 persons, 10 images per person (Fig.1 shows the 10 images of one person). The images are taken at different times, varying lighting slightly, facial expressions (open/closed eyes, smiling/non-smiling) and facial details (glasses/no-glasses). All the images are taken against a dark homogeneous background. The faces are in up-right position of frontal view, with slight left-right out-of-plane rotation. Each image is linearly stretched to the full range of pixel values of [0,255].

Figure 1: Face examples from ORL database.

The set of the 10 images for each person is randomly partitioned into a training subset of 5 images and a test set of the other 5. The training set is then used to learn basis components, and the test set for evaluate. All the compared methods take the same training and test data.

3.2 Learning Basis Components

LNMF, NMF and PCA representations with $25, 36, 49, 64, 81, 100, 121$ basis components are computed from the training set. The matlab package from `http://journalclub.mit.edu` is used for NMF. NMF converges about 5-times faster than LNMF. Fig.2 shows the resulting LNMF and NMF components for subspaces of dimensions 25, 49 and 81. Higher pixel values are in in darker color; the components in each LNMF basis set have been ordered (left-right then top-down) according to the significance value v_{ii}. The NMF bases are as holistic as the PCA basis (eigenfaces) for the training set. We notice the result presented in [1] does not appear so, perhaps because the faces used for producing that result are well aligned. The LNMF procedure learns basis components which not only lead to non-subtractive representations, but also manifest localized features and thus truly parts-based representations. Also, we see that as the dimension (number of components) increases, the features formed in the LNMF components become more localized.
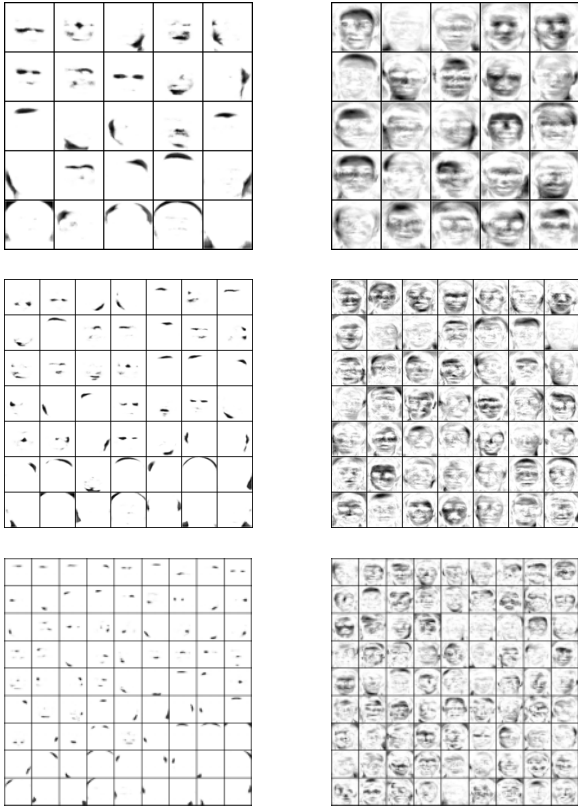
4

Figure 2: LNMF (left) and NMF (right) bases of dimensions 25 (row 1), 49 (row 2) and 81 (row 3). Every basis component is of size 112×92 and the displayed images are re-sized to fit the paper format. The LNMF representation is both parts-based and local, whereas NMF is parts-based but holistic.

3.3 Reconstruction

Fig.3 shows reconstructions in the LNMF, NMF and PCA subspaces of various dimensions for a face image in the test set which corresponds to the one in the middle of row 1 of Fig.1. As the dimension is increased, more details are recovered. We see that while NMF and PCA reconstructions look similar in terms of the smoothness and texture of the reconstructed images, with PCA presenting better reconstruction quality than NMF. Surprisingly the LNMF representation, which is based on more localized features, provides smoother reconstructions than NMF and PCA.

3.4 Face Recognition

The LNMF, NMF and PCA representations are comparatively evaluated for face recognition using the images from the test set. The recognition accuracy, defined as the percentage of correctly recognized faces, is used as the performance measure. Tests are done with varying number of

Figure 3: Reconstructions of the face image in the (left to right) 25, 49, 81 and 121 dimensional (top-down) LNMF, NMF, and PCA subspaces.
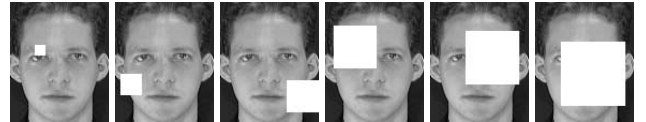

Figure 4: Examples of random occluding patches of sizes (from left to right) 10x10, 20x20, ..., 50x50, 60x60.

basis components, with or without occlusion. The occlusion is simulated in an image by using a white patch of size $s \times s$ with $s \in \{10, 20, \ldots, 60\}$ at a random location; see Fig.4 for examples.

Figs.5 and 6 show recognition accuracy curves under various conditions. Fig.5 compares the three representations in terms of the recognition accuracies versus the number $m \times m$ of basis components for $m \in \{5, 6, \ldots, 10, 11\}$. The LNMF yields the best recognition accuracy, slightly better than PCA whereas the original NMF gives very low accuracy. Fig.6 compares the three representations under varying degrees of occlusion and with varying number of basis components, in terms of the recognition accuracies versus the size $s \times s$ of occluding patch for $s \in \{10, 20, \ldots, 50, 60\}$. As we see, although PCA yields more favorable results than LNMF when the patch size is small, the better stability of the LNMF representation under partial occlusion becomes clear as the patch size increases.

4 Conclusion

In this paper, we have proposed a new method, local non-negative matrix factorization (LNMF), for learning spatially
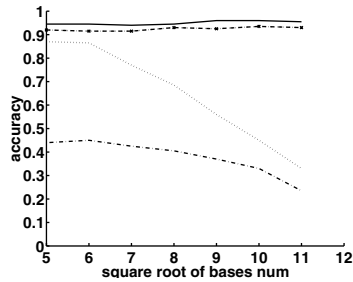
5

Figure 5: Recognition accuracies as function of the number (in 5x5, 6x6, ..., 11x11) of basis components used, for the LNMF (solid) and NMF (dashed) and PCA (dot-dashed) representations.
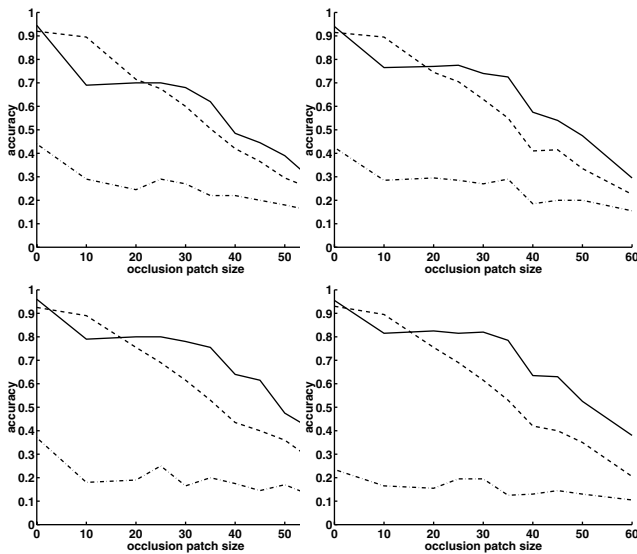

Figure 6: Recognition accuracies versus the size (in 10x10, 20x20, ..., 60x60) of occluding patches, with 25, 49, 81, 121 basis components (left-right, then top-down), for the LNMF (solid) and NMF (dashed) and PCA (dot-dashed) representations.

localized, part-based subspace representation of visual patterns. The work is aimed to learn localized of features in NMF basis components suitable for tasks such as face recognition. An algorithms is presented for the learning and its convergence proved. Experimental results have shown that we have achieved our objectives: LNMF derives bases which are better suited for a localized representation than PCA and NMF, and leads to better recognition results than the existing methods.

The LNMF and NMF learning algorithms are local minimizers. They give different basis components from different initial conditions. We will investigate how this affects the recognition rate. Further future work includes the following topics. The first is to develop algorithms for faster convergence and better solution in terms of minimizing the objective function. The second is to investigate the ability of the model to generalize, *i.e.* how the constraints, the non-negativity and others, are satisfied for data not seen in the training set.The third is to compare with other methods for learning spatially localized features such as LFA [9] and ICA [12].

References

[1] D. D. Lee and H. S. Seung, ""Learning the parts of objects by non-negative matrix factorization"," *Nature*, vol. 401, pp. 788–791, 1999.

[2] L. Sirovich and M. Kirby, ""Low-dimensional procedure for the characterization of human faces"," *Journal of the Optical Society of America A*, vol. 4, no. 3, pp. 519–524, March 1987.

[3] M. Kirby and L. Sirovich, ""Application of the Karhunen-Loeve procedure for the characterization of human faces"," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, January 1990.

[4] Matthew A. Turk and Alex P. Pentland, ""Face recognition using eigenfaces."," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Hawaii, June 1991, pp. 586–591.

[5] David Beymer, Amnon Shashua, and Tomaso Poggio, ""Example based image analysis and synthesis"," A. I. Memo 1431, MIT, 1993.

[6] A. P. Pentland, B. Moghaddam, and T. Starner, ""View-based and modular eigenspaces for face recognition"," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994, pp. 84–91.

[7] H. Murase and S. K. Nayar, ""Visual learning and recognition of 3-D objects from appearance"," *International Journal of Computer Vision*, vol. 14, pp. 5–24, 1995.

[8] K. Fukunaga, *Introduction to statistical pattern recognition*, Academic Press, Boston, 2 edition, 1990.

[9] P. Penev and J. Atick, ""Local feature analysis: A general statistical theory for object representation"," *Neural Systems*, vol. 7, no. 3, pp. 477–500, 1996.

[10] C. Jutten and J. Herault, ""Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture"," *Signal Processing*, vol. 24, pp. 1–10, 1991.

[11] P. Comon, ""Independent component analysis - a new concept?"," *Signal Processing*, vol. 36, pp. 287–314, 1994.

[12] A. J. Bell and T. J. Sejnowski, ""The 'independent components' of natural scenes are edge filters"," *Vision Research*, vol. 37, pp. 3327–3338, 1997.

[13] D. D. Lee and H. S. Seung, ""Algorithms for non-negative matrix factorization"," in *Proceedings of Neural Information Processing Systems*, 2000.

[14] A. P. Dempster, N. M. Laird, and D. B. Bubin, ""Maximum likelihood from imcomplete data via EM algorithm"," *Journal of the Royal Statistical Society, Series B*, vol. 39, pp. 1–38, 1977.